



**Europäisches
Patentamt**

**European
Patent Office**

**Office européen
des brevets**

REC'D 09 JUL 2003

WIPO PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-
gen stimmen mit der
ursprünglich eingereichten
Fassung der auf dem näch-
sten Blatt bezeichneten
europäischen Patentanmel-
dung überein.

The attached documents
are exact copies of the
European patent application
described on the following
page, as originally filed.

Les documents fixés à
cette attestation sont
conformes à la version
initialement déposée de
la demande de brevet
européen spécifiée à la
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02100629.1

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk



Anmeldung Nr:
Application no.: 02100629.1
Demande no:

Anmeldetag:
Date of filing: 03.06.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

DSM N.V.
Het Overloon 1
6411 TE Heerlen
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se référer à la description.)

NOVEL PECTINASE PEC43 AND USES THEREOF

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12N9/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR

NOVEL PECTINASE PEC43 AND USES THEREOF

Field of the invention

5 The invention relates to newly identified polynucleotide sequences comprising genes that encode novel pectinases isolated from *Aspergillus niger*. The invention features the full length nucleotide sequence of the novel gene, the cDNA sequence comprising the full length coding sequence of the novel pectinase as well as the amino acid sequence of the full-length
10 functional protein and functional equivalents thereof. The invention also relates to methods of using these enzymes in industrial processes and methods of diagnosing fungal infections. Also included in the invention are cells transformed with a polynucleotide according to the invention and cells wherein an pectinase according to the invention is genetically modified to enhance or reduce its activity
15 and/or level of expression.

Background of the invention

20 Pectin polymers are important constituents of plant primary cell walls. They are composed of chains of 1,4-linked alpha-D-galacturonic acid and methylated derivatives thereof. Enzymes that are able to degrade the above-defined pectin polymers are called pectinases. Degradation in this respect means that at least one sugar residue or ester group has been removed from the pectin molecule. Pectinases such as polygalacturonase, pectin methylesterase, pectin
25 lyase or pectate lyase are important for the food industry, primarily in the art of fruit and vegetable processing such as fruit juice production or wine making, where their ability to catalyse the degradation of the backbone of the pectin polymer is utilised. Another application of pectinases, in particular pectin methyl esterase, is the firming of fruit and vegetables.

30 The seeds of leguminosae like soy beans and other types of peas also contain pectin. This pectin, however, is quite different from pectins described for fruits like apple berries etc. A large part of soybean pectin is water-insoluble. In addition soybean pectin consists mainly of xylogalacturonan and rhamnogalacturonan and contains only minor amounts of homogalacturonan. The
35 CDTA-soluble pectic substances from soybean meal are composed of

rhamnogalacturonan and xylogalacturonan but not homogalacturonan. Due to these structural differences soybean pectin is much more difficult to degrade by pectinases as compared to fruit pectin. Even the use of technical multi-enzyme preparations was not sufficient to degrade soybean pectin structures (Huisman, MMH; Schols, HA; Voragen, AGJ, **1999**, Enzymatic degradation of cell wall polysaccharides from soybean meal, Carbohydrate Polymers 38, 299-307, and: Ouhida, I; Pérez, JF; Gasa, J; **2002**, Soybean (glycine max) cell wall composition and availability to feed enzymes, J. Agric. Food Chem 50, 1933-1938).

An assortment of different pectin degrading enzymes is known to be present in various microorganisms such as *Aspergillus niger*.

The following is a non-exhaustive list of pectin-acting enzymes:

Enzyme	E.C. Number
exopolygalacturonase	3.2.1.67
endopolygalacturonase	3.2.1.15
pectin lyase	4.2.2.10
pectate lyase	4.2.2.2
rhamnogalacturonan hydrolase	3.2.1.x
rhamnogalacturonan lyase	4.2.2.x
rhamnogalacturonan acetylcetase	3.1.1.x
RG rhamnohydrolase	3.2.1.x
RG galacturonohydrolase	3.2.1.x
xylogalacturonan hydrolase	3.2.1.x
pectin acetylcetase	3.1.1.x
pectin methylesterase	3.1.1.11
endo-arabinase	3.2.1.99
beta-arabinofuranosidase	3.2.1.55
beta-1,4-galactanase	3.2.1.89
beta-1,3-galactanase	3.2.1.90
beta-galactosidase	3.2.1.23
alpha-galactosidase	3.2.1.22

feruloyl acetyl esterase	3.1.1.x
alpha-fucosidase	3.2.1.51
(beta-fucosidase)	3.2.1.38
beta-apiosidase	3.2.1.x
alpha-rhamnosidase	3.2.1.40
beta-rhamnosidase	3.2.1.43
alpha-arabinopyranosidase	3.2.1.x
beta-glucuronidase	3.2.1.31
(alpha-glucuronidase)	3.2.1.139
beta-xylosidase	3.2.1.37
(alpha-xylosidase)	3.2.1.x

Pectin methylesterase catalyses the removal of methanol from pectin, resulting in the formation of pectic acid (polygalacturonic acid). Pectate lyase cleaves glycosidic bonds in polygalacturonic acid by beta-elimination, pectin
5 lyase cleaves the glycosidic bonds of highly methylated pectins by beta-elimination, and polygalacturonase hydrolyses the glycosidic linkages in the polygalacturonic acid chain.

Commercially available pectinases are actually a mixture of enzymes, which, along with other enzymes such as cellulases, are used in the
10 fruit industry to help extract, clarify and modify fruit juices.

It is a disadvantage of these mixtures that they hardly ever contain the optimal mix of enzymes to treat a particular fruit or juice. Optimal pectin degradation depends on many factors, including the kind of fruit to be treated, the season in which the fruit is harvested, the ripeness of the fruit, and
15 many more. A well-known example is the release of unwanted methanol and methylated pectin by the action of pectinmethylesterases and/or exo-polygalacturonases when present in too high concentrations. It is therefore an object of the present invention to provide isolated (recombinant) pectinases. The availability of isolated enzymes allows a cocktail of enzymes to be prepared
20 which is optimized for the digestion of pectin from a particular fruit or juice.

Pectinases have been cloned in a variety of different microorganisms. Molecular cloning of pectinases in fungi has also been described. The DNA and deduced amino acid sequences of pectinases from *Aspergillus oryzae*, *Aspergillus kawachii* and *Emericella nidulans* are known.

25 However, there is still a need for other pectinases with different

properties so that fruit processing may be optimized.

One of the disadvantages of the currently available enzymes is that they have pH optima above pH 4.0. Many fruit juices are more acidic and a pH optimum below 3.8 would be desirable. Fruit juices may become as acidic as pH 2.0 and it is therefore an object of the present invention to provide pectinases with a more acidic pH optimum. Pectinases with a pH optimum below 3.8 would be especially advantageous.

Fruit processing is also often performed at more extreme temperatures. Use of pectinases with lower temperature optima would improve yield and quality of the juice because more aromas would be released. The currently available enzymes have a very low efficiency at temperatures below 15 °C. Also, at high temperatures the currently available enzymes are easily inactivated. This is a disadvantage for those fruit juices that have to be pasteurized.

The enzymes according to the invention provide a better temperature stability and can withstand more extreme temperatures in comparison with pectinases according to the prior art.

Also, the specificity of current enzymes leaves to be desired. Fruit juice that needs to be filtrated after or during processing tends to clot the filters after some time. This is due to the inability of certain enzymes to properly digest fruit polysaccharides which causes the fouling of the filters.

The present invention addresses at least one if not all of the above problems.

Object of the invention

It is an object of the invention to provide novel polynucleotides encoding novel pectinases with improved properties. A further object is to provide naturally and recombinantly produced pectinases as well as recombinant strains producing these. Also fusion polypeptides are part of the invention as well as methods of making and using the polynucleotides and polypeptides according to the invention.

Summary of the invention

The invention provides for novel polynucleotides encoding novel pectinases.

More in particular, the invention provides for polynucleotides having a nucleotide sequence that hybridises preferably under highly stringent

conditions to a sequence according to SEQ ID NO: 1 or SEQ ID NO: 2.

Consequently, the invention provides nucleic acids that are more than 40% such as about 60%, preferably 65%, more preferably 70%, even more preferably 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% homologous to the sequences

5 according to SEQ ID NO: 1 or SEQ ID NO: 2.

In a more preferred embodiment the invention provides for such an isolated polynucleotide obtainable from a filamentous fungus, in particular *A. niger* is preferred.

10 In one embodiment, the invention provides for an isolated polynucleotide comprising a nucleic acid sequence encoding a polypeptide with an amino acid sequence as shown in SEQ ID NO: 3 or functional equivalents thereof.

In a further preferred embodiment, the invention provides an isolated polynucleotide encoding at least one functional domain of a polypeptide according to SEQ ID NO: 3 or functional equivalents thereof.

15 In a preferred embodiment the invention provides an pectinase gene according to SEQ ID NO: 1. In another aspect the invention provides a polynucleotide, preferably a cDNA encoding an *A. niger* pectinase whose amino acid sequence is shown in SEQ ID NO: 3 or variants or fragments of that polypeptide. In a preferred embodiment the cDNA has a sequence according to SEQ ID NO: 2 or functional equivalents thereof.

20 In an even further preferred embodiment, the invention provides for a polynucleotide comprising the coding sequence of the polynucleotides according to the invention, preferred is the polynucleotide sequence of SEQ ID NO: 2.

25 The invention also relates to vectors comprising a polynucleotide sequence according to the invention and primers, probes and fragments that may be used to amplify or detect the DNA according to the invention.

30 In a further preferred embodiment, a vector is provided wherein the polynucleotide sequence according to the invention is functionally linked with regulatory sequences suitable for expression of the encoded amino acid sequence in a suitable host cell, such as *A. niger* or *A. oryzae*. The invention also provides methods for preparing polynucleotides and vectors according to the invention.

35 The invention also relates to recombinantly produced host cells

that contain heterologous or homologous polynucleotides according to the invention.

In another embodiment, the invention provides recombinant host cells wherein the expression of an pectinase according to the invention is
5 significantly increased or wherein the activity of the pectinase is increased.

In another embodiment the invention provides for a recombinantly produced host cell that contains heterologous or homologous DNA according to the invention and wherein the cell is capable of producing a functional pectinase according to the invention, preferably a cell capable of over-
10 expressing the pectinase according to the invention, for example an *Aspergillus* strain comprising an increased copy number of a gene or cDNA according to the invention.

In yet another aspect of the invention, a purified polypeptide is provided. The polypeptides according to the invention include the polypeptides encoded by the polynucleotides according to the invention. Especially preferred is
15 a polypeptide according to SEQ ID NO: 3 or functional equivalents thereof.

Fusion proteins comprising a polypeptide according to the invention are also within the scope of the invention. The invention also provides methods of making the polypeptides according to the invention.

20 The invention also relates to the use of the pectinase according to the invention in any industrial process as described herein

Detailed description of the invention

25 Polynucleotides

The present invention provides polynucleotides encoding an pectinase, tentatively called PEC43, having an amino acid sequence according to SEQ ID NO: 3 or functional equivalents thereof. The sequence of the gene encoding PEC43 was determined by sequencing a genomic clone obtained from
30 *Aspergillus niger*. The invention provides polynucleotide sequences comprising the gene encoding the PEC43 pectinase as well as its complete cDNA sequence and its coding sequence. Accordingly, the invention relates to an isolated polynucleotide comprising the nucleotide sequence according to SEQ ID NO: 1 or SEQ ID NO: 2 or functional equivalents thereof.

35 More in particular, the invention relates to an isolated polynucleotide hybridisable under stringent conditions, preferably under highly stringent conditions, to a polynucleotide according to SEQ ID NO: 1 or SEQ ID

NO: 2. Advantageously, such polynucleotides may be obtained from filamentous fungi, in particular from *Aspergillus niger*. More specifically, the invention relates to an isolated polynucleotide having a nucleotide sequence according to SEQ ID NO: 1 or SEQ ID NO: 2.

5 The invention also relates to an isolated polynucleotide encoding at least one functional domain of a polypeptide according to SEQ ID NO: 3 or functional equivalents thereof.

10 As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which may be isolated from chromosomal DNA, which include an open reading frame encoding a protein, e.g. an *A. niger* pectinase. A gene may include coding sequences, non-coding sequences, introns and regulatory sequences. Moreover, a gene refers to an isolated nucleic acid molecule as defined herein.

15 A nucleic acid molecule of the present invention, such as a nucleic acid molecule having the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO: 2 or a functional equivalent thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, using all or portion of the nucleic acid sequence of SEQ ID NO: 1 or the nucleotide sequence of SEQ ID NO: 2 as a hybridization probe, nucleic acid molecules
20 according to the invention can be isolated using standard hybridization and cloning techniques (e. g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; 1989).

25 Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO: 1 or SEQ ID NO: 2 can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence information contained in SEQ ID NO:1 or SEQ ID NO: 2.

30 A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis.

35 Furthermore, oligonucleotides corresponding to or hybridisable to nucleotide sequences according to the invention can be prepared by standard synthetic techniques, e. g., using an automated DNA synthesizer.

 In a preferred embodiment, an isolated nucleic acid molecule of

the invention comprises the nucleotide sequence shown in SEQ ID NO: 2. The sequence of SEQ ID NO: 2 corresponds to the coding region of the A. niger PEC43 cDNA. This cDNA comprises sequences encoding the A. niger PEC43 polypeptide according to SEQ ID NO: 3.

5 In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO: 2 or a functional equivalent of these nucleotide sequences.

10 A nucleic acid molecule which is complementary to another nucleotide sequence is one which is sufficiently complementary to the other nucleotide sequence such that it can hybridize to the other nucleotide sequence thereby forming a stable duplex.

15 One aspect of the invention pertains to isolated nucleic acid molecules that encode a polypeptide of the invention or a functional equivalent thereof such as a biologically active fragment or domain, as well as nucleic acid molecules sufficient for use as hybridisation probes to identify nucleic acid molecules encoding a polypeptide of the invention and fragments of such nucleic acid molecules suitable for use as PCR primers for the amplification or mutation of nucleic acid molecules.

20 An "isolated polynucleotide" or "isolated nucleic acid" is a DNA or RNA that is not immediately contiguous with both of the coding sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally occurring genome of the organism from which it is derived. Thus, in one embodiment, an isolated nucleic acid includes some or all of the 5' non-coding
25 (e.g., promotor) sequences that are immediately contiguous to the coding sequence. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or
30 restriction endonuclease treatment) independent of other sequences. It also includes a recombinant DNA that is part of a hybrid gene encoding an additional polypeptide that is substantially free of cellular material, viral material, or culture medium (when produced by recombinant DNA techniques), or chemical precursors or other chemicals (when chemically synthesized). Moreover, an
35 "isolated nucleic acid fragment" is a nucleic acid fragment that is not naturally occurring as a fragment and would not be found in the natural state.

As used herein, the terms "polynucleotide" or "nucleic acid molecule" are intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. The nucleic acid may be synthesized using oligonucleotide analogs or derivatives (e.g., inosine or phosphorothioate nucleotides). Such oligonucleotides can be used, for example, to prepare nucleic acids that have altered base-pairing abilities or increased resistance to nucleases.

Another embodiment of the invention provides an isolated nucleic acid molecule which is antisense to a PEC43 nucleic acid molecule, e.g., the coding strand of a PEC43 nucleic acid molecule. Also included within the scope of the invention are the complement strands of the nucleic acid molecules described herein.

Sequencing errors

The sequence information as provided herein should not be so narrowly construed as to require inclusion of erroneously identified bases. The specific sequences disclosed herein can be readily used to isolate the complete gene from filamentous fungi, in particular *A. niger* which in turn can easily be subjected to further sequence analyses thereby identifying sequencing errors.

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of a DNA sequence determined as above. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely

different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

The person skilled in the art is capable of identifying such erroneously identified bases and knows how to correct for such errors.

5

Nucleic acid fragments, probes and primers

A nucleic acid molecule according to the invention may comprise only a portion or a fragment of the nucleic acid sequence shown in SEQ
10 ID NO:1 or SEQ ID NO:2, for example a fragment which can be used as a probe or primer or a fragment encoding a portion of a PEC43 protein. The nucleotide sequence determined from the cloning of the PEC43 gene and cDNA allows for the generation of probes and primers designed for use in identifying and/or cloning other PEC43 family members, as well as PEC43 homologues from other
15 species. The probe/primer typically comprises substantially purified oligonucleotide which typically comprises a region of nucleotide sequence that hybridizes preferably under highly stringent conditions to at least about 12 or 15, preferably about 18 or 20, preferably about 22 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 or more consecutive nucleotides of a nucleotide
20 sequence shown in SEQ ID NO:1 or SEQ ID NO: 2 or of a functional equivalent thereof.

Probes based on the PEC43 nucleotide sequences can be used to detect transcripts or genomic PEC43 sequences encoding the same or homologous proteins for instance in other organisms. In preferred embodiments,
25 the probe further comprises a label group attached thereto, e.g., the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme cofactor. Such probes can also be used as part of a diagnostic test kit for identifying cells which express a PEC43 protein.

Identity & homology

The terms "homology" or "percent identity" are used interchangeably herein. For the purpose of this invention, it is defined here that in order to determine the percent identity of two amino acid sequences or of two
35 nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid

sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = number of identical positions/total number of positions (i.e. overlapping positions) x 100). Preferably, the two sequences are the same length.

The skilled person will be aware of the fact that several different computer programmes are available to determine the homology between two sequences. For instance, a comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (J. Mol. Biol. (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. The skilled person will appreciate that all these different parameters will yield slightly different results but that the overall percentage identity of two sequences is not significantly altered when using different algorithms.

In yet another embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity two amino acid or nucleotide sequence is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989) which has been incorporated into the ALIGN program (version 2.0) (available at: <http://vega.igh.cnrs.fr/bin/align-guess.cgi>) using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) J. Mol. Biol. 215:403—10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength =

12 to obtain nucleotide sequences homologous to PEC43 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to PEC43 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) Nucleic Acids Res. 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

10 **Hybridisation**

As used herein, the term "hybridizing" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 50%, at least about 60%, at least about 70%, more preferably at least about 80%, even more preferably at least about 85% to 90%, more preferably at least 95% homologous to each other typically remain hybridized to each other.

A preferred, non-limiting example of such hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45 °C, followed by one or more washes in 1 X SSC, 0.1 % SDS at 50 °C, preferably at 55 °C, preferably at 60 °C and even more preferably at 65 °C.

Highly stringent conditions include, for example, hybridizing at 68 °C in 5x SSC/5x Denhardt's solution / 1.0% SDS and washing in 0.2x SSC/0.1% SDS at room temperature. Alternatively, washing may be performed at 42 °C.

The skilled artisan will know which conditions to apply for stringent and highly stringent hybridisation conditions. Additional guidance regarding such conditions is readily available in the art, for example, in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N.Y.; and Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N.Y.).

Of course, a polynucleotide which hybridizes only to a poly A sequence (such as the 3' terminal poly(A) tract of mRNAs), or to a complementary stretch of T (or U) residues, would not be included in a polynucleotide of the invention used to specifically hybridize to a portion of a nucleic acid of the invention, since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-standed cDNA clone).

Obtaining full length DNA from other organisms

5 In a typical approach, cDNA libraries constructed from other organisms, e.g. filamentous fungi, in particular from the species *Aspergillus* can be screened.

For example, *Aspergillus* strains can be screened for homologous PEC43 polynucleotides by Northern blot analysis. Upon detection of transcripts homologous to polynucleotides according to the invention, cDNA
10 libraries can be constructed from RNA isolated from the appropriate strain, utilizing standard techniques well known to those of skill in the art. Alternatively, a total genomic DNA library can be screened using a probe hybridisable to a PEC43 polynucleotide according to the invention.

Homologous gene sequences can be isolated, for example, by
15 performing PCR using two degenerate oligonucleotide primer pools designed on the basis of nucleotide sequences as taught herein.

The template for the reaction can be cDNA obtained by reverse transcription of mRNA prepared from strains known or suspected to express a polynucleotide according to the invention. The PCR product can be subcloned and
20 sequenced to ensure that the amplified sequences represent the sequences of a new PEC43 nucleic acid sequence, or a functional equivalent thereof.

The PCR fragment can then be used to isolate a full length cDNA clone by a variety of known methods. For example, the amplified fragment can be labeled and used to screen a bacteriophage or cosmid cDNA library.
25 Alternatively, the labeled fragment can be used to screen a genomic library.

PCR technology also can be used to isolate full length cDNA sequences from other organisms. For example, RNA can be isolated, following standard procedures, from an appropriate cellular or tissue source. A reverse transcription reaction can be performed on the RNA using an oligonucleotide
30 primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis.

The resulting RNA/DNA hybrid can then be "tailed" (e.g., with guanines) using a standard terminal transferase reaction, the hybrid can be digested with RNase H, and second strand synthesis can then be primed (e.g.,
35 with a poly-C primer). Thus, cDNA sequences upstream of the amplified fragment can easily be isolated. For a review of useful cloning strategies, see e.g., Sambrook et al., supra; and Ausubel et al., supra.

Vectors

Another aspect of the invention pertains to vectors, preferably
5 expression vectors, containing a nucleic acid encoding a PEC43 protein or a
functional equivalent thereof. As used herein, the term "vector" refers to a nucleic
acid molecule capable of transporting another nucleic acid to which it has been
linked. One type of vector is a "plasmid", which refers to a circular double stranded
10 DNA loop into which additional DNA segments can be ligated. Another type of
vector is a viral vector, wherein additional DNA segments can be ligated into the
viral genome. Certain vectors are capable of autonomous replication in a host cell
into which they are introduced (e.g., bacterial vectors having a bacterial origin of
replication and episomal mammalian vectors). Other vectors (e.g., non-episomal
15 mammalian vectors) are integrated into the genome of a host cell upon
introduction into the host cell, and thereby are replicated along with the host
genome. Moreover, certain vectors are capable of directing the expression of
genes to which they are operatively linked. Such vectors are referred to herein as
"expression vectors". In general, expression vectors of utility in recombinant DNA
20 techniques are often in the form of plasmids. The terms "plasmid" and "vector" can
be used interchangeably herein as the plasmid is the most commonly used form
of vector. However, the invention is intended to include such other forms of
expression vectors, such as viral vectors (e.g., replication defective retroviruses,
adenoviruses and adeno-associated viruses), which serve equivalent functions:

The recombinant expression vectors of the invention comprise a
25 nucleic acid of the invention in a form suitable for expression of the nucleic acid in
a host cell, which means that the recombinant expression vector includes one or
more regulatory sequences, selected on the basis of the host cells to be used for
expression, which is operatively linked to the nucleic acid sequence to be
expressed. Within a recombinant expression vector, "operatively linked" is
30 intended to mean that the nucleotide sequence of interest is linked to the
regulatory sequence(s) in a manner which allows for expression of the nucleotide
sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when
the vector is introduced into the host cell). The term "regulatory sequence" is
intended to include promoters, enhancers and other expression control elements
35 (e.g., polyadenylation signal). Such regulatory sequences are described, for
example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185,
Academic Press, San Diego, CA (1990). Regulatory sequences include those

which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in a certain host cell (e.g. tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, encoded by nucleic acids as described herein (e.g. PEC43 proteins, mutant forms of PEC43 proteins, fragments, variants or functional equivalents thereof, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of PEC43 proteins in prokaryotic or eukaryotic cells. For example, PEC43 proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression vectors useful in the present invention include chromosomal-, episomal- and virus-derived vectors e.g., vectors derived from bacterial plasmids, bacteriophage, yeast episome, yeast chromosomal elements, viruses such as baculoviruses, papova viruses, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids.

The DNA insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli* lac, trp and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled person. In a specific embodiment, promoters are preferred that are capable of directing a high expression level of pectinases in filamentous fungi. Such promoters are known in the art. The expression constructs may contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation.

The coding portion of the mature transcripts expressed by the constructs will include a translation initiating AUG at the beginning and a termination codon

appropriately positioned at the end of the polypeptide to be translated.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, transduction, infection, lipofection, cationic lipid-mediated transfection or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual, 2nd, ed.* Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), Davis et al., *Basic Methods in Molecular Biology* (1986) and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a PEC43 protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g. cells that have incorporated the selectable marker gene will survive, while the other cells die).

Expression of proteins in prokaryotes is often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, e.g. to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and

enterokinase.

As indicated, the expression vectors will preferably contain selectable markers. Such markers include dihydrofolate reductase or neomycin resistance for eukarotic cell culture and tetracycline or ampicillin resistance for culturing in *E. coli* and other bacteria. Representative examples of appropriate host include bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium*; fungal cells, such as yeast; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9; animal cells such as CHO, COS and Bowes melanoma; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria are pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16A, pNH18A, pNH46A, available from Stratagene; and pTrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pZT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Among known bacterial promoters for use in the present invention include *E. coli* lacI and lacZ promoters, the T3 and T7 promoters, the gpt promoter, the lambda PR, PL promoters and the trp promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous sarcoma virus ("RSV"), and metallothionein promoters, such as the mouse metallothionein-I promoter.

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act to increase transcriptional activity of a promoter in a given host cell-type. Examples of enhancers include the SV40 enhancer, which is located on the late side of the replication origin at bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signal may be incorporated into the expressed polypeptide. The signals may be endogenous to the polypeptide or they may be heterologous signals.

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification.

Polypeptides according to the invention

The invention provides an isolated polypeptide having the amino acid sequence according to SEQ ID NO: 3, an amino acid sequence obtainable by expressing the polynucleotide of SEQ ID NO: 1 in an appropriate host, as well as an amino acid sequence obtainable by expressing the polynucleotide sequences of SEQ ID NO: 2 in an appropriate host. Also, a peptide or polypeptide comprising a functional equivalent of the above polypeptides is comprised within the present invention. The above polypeptides are collectively comprised in the term "polypeptides according to the invention"

The terms "peptide" and "oligopeptide" are considered synonymous (as is commonly recognized) and each term can be used interchangeably as the context requires to indicate a chain of at least two amino acids coupled by peptidyl linkages. The word "polypeptide" is used herein for chains containing more than seven amino acid residues. All oligopeptide and polypeptide formulas or sequences herein are written from left to right and in the direction from amino terminus to carboxy terminus. The one-letter code of amino acids used herein is commonly known in the art and can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual*, 2nd, ed. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989)

By "isolated" polypeptide or protein is intended a polypeptide or protein removed from its native environment. For example, recombinantly produced polypeptides and proteins expressed in host cells are considered isolated for the purpose of the invention as are native or recombinant polypeptides which have been substantially purified by any suitable technique such as, for example, the single-step purification method disclosed in Smith and Johnson, Gene 67:31-40 (1988).

The PEC43 pectinase according to the invention can be recovered and purified from recombinant cell cultures by well-known methods

including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

Protein fragments

The invention also features biologically active fragments of the polypeptides according to the invention.

Biologically active fragments of a polypeptide of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the PEC43 protein (e.g., the amino acid sequence of SEQ ID NO: 3), which include fewer amino acids than the full length protein, and exhibit at least one biological activity of the corresponding full-length protein. Typically, biologically active fragments comprise a domain or motif with at least one activity of the PEC43 protein. A biologically active fragment of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the biological activities of the native form of a polypeptide of the invention.

The invention also features nucleic acid fragments which encode the above biologically active fragments of the PEC43 protein.

Fusion proteins

The proteins of the present invention or functional equivalents thereof, e.g., biologically active portions thereof, can be operatively linked to a

non-PEC43 polypeptide (e.g., heterologous amino acid sequences) to form fusion proteins. As used herein, a PEC43 "chimeric protein" or "fusion protein" comprises a PEC43 polypeptide operatively linked to a non-PEC43 polypeptide. A "PEC43 polypeptide" refers to a polypeptide having an amino acid sequence
5 corresponding to PEC43, whereas a "non-PEC43 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the PEC43 protein, e.g., a protein which is different from the PEC43 protein and which is derived from the same or a different organism. Within a PEC43 fusion protein the PEC43 polypeptide can correspond
10 to all or a portion of a PEC43 protein. In a preferred embodiment, a PEC43 fusion protein comprises at least one biologically active fragment of a PEC43 protein. In another preferred embodiment, a PEC43 fusion protein comprises at least two biologically active portions of a PEC43 protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the PEC43 polypeptide and the
15 non-PEC43 polypeptide are fused in-frame to each other. The non-PEC43 polypeptide can be fused to the N-terminus or C-terminus of the PEC43 polypeptide.

For example, in one embodiment, the fusion protein is a GST-PEC43 fusion protein in which the PEC43 sequences are fused to the C-terminus
20 of the GST sequences. Such fusion proteins can facilitate the purification of recombinant PEC43. In another embodiment, the fusion protein is a PEC43 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian and Yeast host cells), expression and/or secretion of PEC43 can be increased through use of a heterologous signal sequence.

25 In another example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (*Current Protocols in Molecular Biology*, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase
30 (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (Sambrook et al., *supra*) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

A signal sequence can be used to facilitate secretion and
35 isolation of a protein or polypeptide of the invention. Signal sequences are typically characterized by a core of hydrophobic amino acids which are generally

cleaved from the mature protein during secretion in one or more cleavage events. Such signal peptides contain processing sites that allow cleavage of the signal sequence from the mature proteins as they pass through the secretory pathway. The signal sequence directs secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal sequence is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by art recognized methods. Alternatively, the signal sequence can be linked to the protein of interest using a sequence which facilitates purification, such as with a GST domain. Thus, for instance, the sequence encoding the polypeptide may be fused to a marker sequence, such as a sequence encoding a peptide, which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (Qiagen, Inc.), among others, many of which are commercially available. As described in *Gentz et al, Proc. Natl. Acad. Sci. USA 86:821-824 (1989)*, for instance, hexa-histidine provides for convenient purification of the fusion protein. The HA tag is another peptide useful for purification which corresponds to an epitope derived of influenza hemagglutinin protein, which has been described by Wilson *et al., Cell 37:767 (1984)*, for instance.

Preferably, a PEC43 chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g, a GST polypeptide). A PEC43-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the PEC43

protein.

Functional equivalents

5 The terms "functional equivalents" and "functional variants" are used interchangeably herein. Functional equivalents of PEC43 DNA are isolated DNA fragments that encode a polypeptide that exhibits a particular function of the PEC43 *A. niger* pectinase as defined herein. A functional equivalent of a PEC43 polypeptide according to the invention is a polypeptide that exhibits at least one
10 function of an *A. niger* pectinase as defined herein. Functional equivalents therefore also encompass biologically active fragments.

 Functional protein or polypeptide equivalents may contain only conservative substitutions of one or more amino acids of SEQ ID NO: 3 or substitutions, insertions or deletions of non-essential amino acids. Accordingly, a
15 non-essential amino acid is a residue that can be altered in SEQ ID NO: 3 without substantially altering the biological function. For example, amino acid residues that are conserved among the PEC43 proteins of the present invention, are predicted to be particularly unamenable to alteration. Furthermore, amino acids conserved among the PEC43 proteins according to the present invention and
20 other pectinases are not likely to be amenable to alteration.

 The term "conservative substitution" is intended to mean that a substitution in which the amino acid residue is replaced with an amino acid residue having a similar side chain. These families are known in the art and include amino acids with basic side chains (e.g. lysine, arginine and histidine),
25 acidic side chains (e.g. aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), non-polar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan,
30 histidine).

 Functional nucleic acid equivalents may typically contain silent mutations or mutations that do not alter the biological function of encoded polypeptide. Accordingly, the invention provides nucleic acid molecules encoding PEC43 proteins that contain changes in amino acid residues that are not essential
35 for a particular biological activity. Such PEC43 proteins differ in amino acid sequence from SEQ ID NO: 3 yet retain at least one biological activity. In one embodiment the isolated nucleic acid molecule comprises a nucleotide sequence

encoding a protein, wherein the protein comprises a substantially homologous amino acid sequence of at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homologous to the amino acid sequence shown in SEQ ID NO: 3.

5 For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J.U. et al., Science 247:1306-1310 (1990) wherein the authors indicate that there are two main approaches for studying the tolerance of an amino acid sequence to change. The first method relies on the process of evolution, in which mutations are either accepted or
10 rejected by natural selection. The second approach uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene and selects or screens to identify sequences that maintain functionality. As the authors state, these studies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which changes are likely to be
15 permissive at a certain position of the protein. For example, most buried amino acid residues require non-polar side chains, whereas few features of surface side chains are generally conserved. Other such phenotypically silent substitutions are described in Bowie et al, supra, and the references cited therein.

 An isolated nucleic acid molecule encoding a PEC43 protein
20 homologous to the protein according to SEQ ID NO: 3 can be created by introducing one or more nucleotide substitutions, additions or deletions into the coding nucleotide sequences according to SEQ ID NO: 1 or SEQ ID NO: 2 such that one or more amino acid substitutions, deletions or insertions are introduced into the encoded protein. Such mutations may be introduced by standard
25 techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis.

 The term "functional equivalents" also encompasses orthologues of the A. niger PEC43 protein. Orthologues of the A. niger PEC43 protein are proteins that can be isolated from other strains or species and possess a similar or identical biological activity. Such orthologues can readily be identified
30 as comprising an amino acid sequence that is substantially homologous to SEQ ID NO: 3.

 As defined herein, the term "substantially homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., with similar side chain) amino acids or
35 nucleotides to a second amino acid or nucleotide sequence such that the first and the second amino acid or nucleotide sequences have a common domain. For

example, amino acid or nucleotide sequences which contain a common domain having about 60%, preferably 65%, more preferably 70%, even more preferably 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identity or more are defined herein as sufficiently identical.

5 Also, nucleic acids encoding other PEC43 family members, which thus have a nucleotide sequence that differs from SEQ ID NO: 1 or SEQ ID NO: 2, are within the scope of the invention. Moreover, nucleic acids encoding PEC43 proteins from different species which thus have a nucleotide sequence which differs from SEQ ID NO: 1 or SEQ ID NO: 2 are within the scope of the
10 invention.

 Nucleic acid molecules corresponding to variants (e.g. natural allelic variants) and homologues of the PEC43 DNA of the invention can be isolated based on their homology to the PEC43 nucleic acids disclosed herein using the cDNAs disclosed herein or a suitable fragment thereof, as a
15 hybridisation probe according to standard hybridisation techniques preferably under highly stringent hybridisation conditions.

 In addition to naturally occurring allelic variants of the PEC43 sequence, the skilled person will recognise that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO: 1 or SEQ ID NO: 2 thereby
20 leading to changes in the amino acid sequence of the PEC43 protein without substantially altering the function of the PEC43 protein.

 In another aspect of the invention, improved PEC43 proteins are provided. Improved PEC43 proteins are proteins wherein at least one biological activity is improved. Such proteins may be obtained by randomly introducing
25 mutations along all or part of the PEC43 coding sequence, such as by saturation mutagenesis, and the resulting mutants can be expressed recombinantly and screened for biological activity. For instance, the art provides for standard assays for measuring the enzymatic activity of pectinases and thus improved proteins may easily be selected.

30 In a preferred embodiment the PEC43 protein has an amino acid sequence according to SEQ ID NO: 3. In another embodiment, the PEC43 polypeptide is substantially homologous to the amino acid sequence according to SEQ ID NO: 3 and retains at least one biological activity of a polypeptide according to SEQ ID NO: 3, yet differs in amino acid sequence due to natural
35 variation or mutagenesis as described above.

 In a further preferred embodiment, the PEC43 protein has an

amino acid sequence encoded by an isolated nucleic acid fragment capable of hybridising to a nucleic acid according to SEQ ID NO: 1 or SEQ ID NO: 2, preferably under highly stringent hybridisation conditions.

Accordingly, the PEC43 protein is a protein which comprises an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homologous to the amino acid sequence shown in SEQ ID NO: 3 and retains at least one functional activity of the polypeptide according to SEQ ID NO: 3.

Functional equivalents of a protein according to the invention can also be identified e.g. by screening combinatorial libraries of mutants, e.g. truncation mutants, of the protein of the invention for pectinase activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display). There are a variety of methods that can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang (1983) Tetrahedron 39:3; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477).

In addition, libraries of fragments of the coding sequence of a polypeptide of the invention can be used to generate a variegated population of polypeptides for screening a subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the protein of interest.

Several techniques are known in the art for screening gene

products of combinatorial libraries made by point mutations of truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into
5 replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can
10 be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan (1992) Proc. Natl. Acad. Sci. USA 89:7811-7815; Delgrave et al. (1993) Protein Engineering 6(3):327-331).

In addition to the PEC43 gene sequence shown in SEQ ID NO: 1, it will be apparent for the person skilled in the art that DNA sequence
15 polymorphisms that may lead to changes in the amino acid sequence of the PEC43 protein may exist within a given population. Such genetic polymorphisms may exist in cells from different populations or within a population due to natural allelic variation. Allelic variants may also include functional equivalents.

Fragments of a polynucleotide according to the invention may
20 also comprise polynucleotides not encoding functional polypeptides. Such polynucleotides may function as probes or primers for a PCR reaction.

Nucleic acids according to the invention irrespective of whether they encode functional or non-functional polypeptides, can be used as hybridization probes or polymerase chain reaction (PCR) primers. Uses of the
25 nucleic acid molecules of the present invention that do not encode a polypeptide having a PEC43 activity include, inter alia, (1) isolating the gene encoding the PEC43 protein, or allelic variants thereof from a cDNA library e.g. from other organisms than *A. niger*; (2) in situ hybridization (e.g. FISH) to metaphase chromosomal spreads to provide precise chromosomal location of the PEC43
30 gene as described in Verma et al., Human Chromosomes: a Manual of Basic Techniques, Pergamon Press, New York (1988); (3) Northern blot analysis for detecting expression of PEC43 mRNA in specific tissues and/or cells and 4) probes and primers that can be used as a diagnostic tool to analyse the presence of a nucleic acid hybridisable to the PEC43 probe in a given biological (e.g. tissue)
35 sample.

Also encompassed by the invention is a method of obtaining a

functional equivalent of a PEC43 gene or cDNA. Such a method entails obtaining a labelled probe that includes an isolated nucleic acid which encodes all or a portion of the sequence according to SEQ ID NO: 3 or a variant thereof; screening a nucleic acid fragment library with the labelled probe under conditions that allow hybridisation of the probe to nucleic acid fragments in the library, thereby forming nucleic acid duplexes, and preparing a full-length gene sequence from the nucleic acid fragments in any labelled duplex to obtain a gene related to the PEC43 gene.

5 In one embodiment, a PEC43 nucleic acid of the invention is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 10 97%, 98%, 99%, or more homologous to a nucleic acid sequence shown in SEQ ID NO: 1, SEQ ID NO: 2 or the complement thereof.

In another preferred embodiment a PEC43 polypeptide of the invention is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to the amino acid sequence shown in SEQ ID NO: 3.

15

Host cells

In another embodiment, the invention features cells, e.g., transformed host cells or recombinant host cells that contain a nucleic acid encompassed by the invention. A "transformed cell" or "recombinant cell" is a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a nucleic acid according to the invention. Both prokaryotic and eukaryotic cells are included, e.g., bacteria, fungi, yeast, and the like, especially preferred are cells from filamentous fungi, in particular *Aspergillus niger*.

20

25

A host cell can be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in a specific, desired fashion. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may facilitate optimal functioning of the protein.

30

Various host cells have characteristic and specific mechanisms for post-translational processing and modification of proteins and gene products. Appropriate cell lines or host systems familiar to those of skill in the art of molecular biology and/or microbiology can be chosen to ensure the desired and correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product can

35

be used. Such host cells are well known in the art.

Host cells also include, but are not limited to, mammalian cell lines such as CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, and choroid plexus cell lines.

5 If desired, the polypeptides according to the invention can be produced by a stably-transfected cell line. A number of vectors suitable for stable transfection of mammalian cells are available to the public, methods for constructing such cell lines are also publicly known, e.g., in Ausubel et al. (supra).

10 **Antibodies**

The invention further features antibodies, such as monoclonal or polyclonal antibodies, that specifically bind PEC43 proteins according to the invention.

15 As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')₂ fragments) which are capable of specifically binding to PEC43 protein. Fab and F(ab')₂ fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have
20 less non-specific tissue binding of an intact antibody (Wahl et al., *J. Nucl. Med.* 24:316-325 (1983)). Thus, these fragments are preferred.

The antibodies of the present invention may be prepared by any of a variety of methods. For example, cells expressing the PEC43 protein or an antigenic fragment thereof can be administered to an animal in order to induce the
25 production of sera containing polyclonal antibodies. In a preferred method, a preparation of PEC43 protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present
30 invention are monoclonal antibodies (or PEC43 protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology (Kohler et al., *Nature* 256:495 (1975); Kohler et al., *Eur. J. Immunol.* 6:511 (1976); Hammerling et al., In: *Monoclonal Antibodies and T-Cell Hybridomas*, Elsevier, N.Y., pp. 563-681 (1981)). In general, such procedures
35 involve immunizing an animal (preferably a mouse) with a PEC43 protein antigen or, with a PEC43 protein expressing cell. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell

line may be employed in accordance with the present invention; however, it is preferably to employ the parent myeloma cell line (SP₂O), available from the American Type Culture Collection, Rockville, Maryland. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by
5 limiting dilution as described by Wands *et al.* (*Gastro-enterology* 80:225-232 (1981)). The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the PEC43 protein antigen. In general, the polypeptides can be coupled to a carrier protein, such as KLH, as described in Ausubel *et al.*, supra, mixed with an adjuvant, and injected
10 into a host mammal.

In particular, various host animals can be immunized by injection of a polypeptide of interest. Examples of suitable host animals include rabbits, mice, guinea pigs, and rats. Various adjuvants can be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), adjuvant mineral gels such as aluminum
15 hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*. Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of the immunized animals.
20

Such antibodies can be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD, and any subclass thereof. The hybridomas producing the mAbs of this invention can be cultivated *in vitro* or *in vivo*.

Once produced, polyclonal or monoclonal antibodies are tested
25 for specific recognition of a PEC43 polypeptide or functional equivalent thereof in an immunoassay, such as a Western blot or immunoprecipitation analysis using standard techniques, e.g., as described in Ausubel *et al.*, supra. Antibodies that specifically bind to PEC43 proteins or functional equivalents thereof are useful in the invention. For example, such antibodies can be used in an immunoassay to
30 detect PEC43 in pathogenic or non-pathogenic strains of *Aspergillus* (e.g., in *Aspergillus* extracts).

Preferably, antibodies of the invention are produced using fragments of the PEC43 polypeptides that appear likely to be antigenic, by criteria such as high frequency of charged residues. For example, such fragments may be
35 generated by standard techniques of PCR, and then cloned into the pGEX expression vector (Ausubel *et al.*, supra). Fusion proteins may then be expressed

in *E. coli* and purified using a glutathione agarose affinity matrix as described in Ausubel, et al., supra. If desired, several (e.g., two or three) fusions can be generated for each protein, and each fusion can be injected into at least two rabbits. Antisera can be raised by injections in a series, typically including at least
5 three booster injections. Typically, the antisera are checked for their ability to immunoprecipitate a recombinant PEC43 polypeptide or functional equivalents thereof whereas unrelated proteins may serve as a control for the specificity of the immune reaction.

Alternatively, techniques described for the production of single
10 chain antibodies (U.S. Patent 4,946,778 and 4,704,692) can be adapted to produce single chain antibodies against a PEC43 polypeptide or functional equivalents thereof. Kits for generating and screening phage display libraries are commercially available e.g. from Pharmacia.

Additionally, examples of methods and reagents particularly
15 amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223, 409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 20791; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO
20 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J.* 12:725-734.

Polyclonal and monoclonal antibodies that specifically bind
25 PEC43 polypeptides of functional equivalents thereof can be used, for example, to detect expression of a PEC43 gene or a functional equivalent thereof e.g. in another strain of *Aspergillus*. For example, PEC43 polypeptide can be readily detected in conventional immunoassays of *Aspergillus* cells or extracts. Examples of suitable assays include, without limitation, Western blotting, ELISAs,
30 radioimmune assays, and the like.

By "specifically binds" is meant that an antibody recognizes and binds a particular antigen, e.g., a PEC43 polypeptide, but does not substantially recognize and bind other unrelated molecules in a sample.

Antibodies can be purified, for example, by affinity
35 chromatography methods in which the polypeptide antigen is immobilized on a resin.

An antibody directed against a polypeptide of the invention (e.g., monoclonal antibody) can be used to isolate the polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, such an antibody can be used to detect the protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. The antibodies can also be used diagnostically to monitor protein levels in cells or tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen or in the diagnosis of Aspergillosis..

Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive materials include ^{125}I , ^{131}I , ^{35}S or ^3H .

Preferred epitopes encompassed by the antigenic peptide are regions that are located on the surface of the protein, e.g., hydrophilic regions. Hydrophobicity plots of the proteins of the invention can be used to identify hydrophilic regions.

The antigenic peptide of a protein of the invention comprises at least 7 (preferably 10, 15, 20, or 30) contiguous amino acid residues of the amino acid sequence of SEQ ID NO: 3 and encompasses an epitope of the protein such that an antibody raised against the peptide forms a specific immune complex with the protein.

Preferred epitopes encompassed by the antigenic peptide are regions of PEC43 that are located on the surface of the protein, e.g., hydrophilic regions, hydrophobic regions, alpha regions, beta regions, coil regions, turn regions and flexible regions.

Immunoassays

Qualitative or quantitative determination of a polypeptide

according to the present invention in a biological sample can occur using any art-known method. Antibody-based techniques provide special advantages for assaying specific polypeptide levels in a biological sample.

5 In these, the specific recognition is provided by the primary antibody (polyclonal or monoclonal) but the secondary detection system can utilize fluorescent, enzyme, or other conjugated secondary antibodies. As a result, an immunocomplex is obtained.

Accordingly, the invention provides a method for diagnosing whether a certain organism is infected with *Aspergillus* comprising the steps of:

- 10
- Isolating a biological sample from said organism suspected to be infected with *Aspergillus*,
 - reacting said biological sample with an antibody according to the invention,
 - determining whether immunocomplexes are formed.

15 Tissues can also be extracted, e.g., with urea and neutral detergent, for the liberation of protein for Western-blot or dot/slot assay. This technique can also be applied to body fluids.

Other antibody-based methods useful for detecting PEC43 gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). For example, PEC43-specific
20 monoclonal antibodies can be used both as an immunoabsorbent and as an enzyme-labeled probe to detect and quantify the PEC43 protein. The amount of PEC43 protein present in the sample can be calculated by reference to the amount present in a standard preparation using a linear regression computer algorithm. In another ELISA assay, two distinct specific monoclonal antibodies can
25 be used to detect PEC43 protein in a biological fluid. In this assay, one of the antibodies is used as the immuno-absorbent and the other as the enzyme-labeled probe.

The above techniques may be conducted essentially as a "one-step" or "two-step" assay. The "one-step" assay involves contacting PEC43
30 protein with immobilized antibody and, without washing, contacting the mixture with the labeled antibody. The "two-step" assay involves washing before contacting the mixture with the labeled antibody. Other conventional methods may also be employed as suitable. It is usually desirable to immobilize one component of the assay system on a support, thereby allowing other components of the
35 system to be brought into contact with the component and readily removed from the sample.

Suitable enzyme labels include, for example, those from the oxidase group, which catalyze the production of hydrogen peroxide by reacting with substrate. Activity of an oxidase label may be assayed by measuring the concentration of hydrogen peroxide formed by the enzyme-labelled antibody/substrate reaction.

Besides enzymes, other suitable labels include radioisotopes, such as iodine (^{125}I , ^{127}I), carbon (^{14}C), sulphur (^{35}S), tritium (^3H), indium (^{112}In), and technetium ($^{99\text{m}}\text{Tc}$), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

Specific binding of a test compound to a PEC43 polypeptide can be detected, for example, in vitro by reversibly or irreversibly immobilizing the PEC43 polypeptide on a substrate, e.g., the surface of a well of a 96-well polystyrene microtitre plate. Methods for immobilizing polypeptides and other small molecules are well known in the art. For example, the microtitre plates can be coated with a PEC43 polypeptide by adding the polypeptide in a solution (typically, at a concentration of 0.05 to 1 mg/ml in a volume of 1-100 μl) to each well, and incubating the plates at room temperature to 37 $^{\circ}\text{C}$ for 0.1 to 36 hours. Polypeptides that are not bound to the plate can be removed by shaking the excess solution from the plate, and then washing the plate (once or repeatedly) with water or a buffer. Typically, the polypeptide is contained in water or a buffer. The plate is then washed with a buffer that lacks the bound polypeptide. To block the free protein-binding sites on the plates, the plates are blocked with a protein that is unrelated to the bound polypeptide. For example, 300 μl of bovine serum albumin (BSA) at a concentration of 2 mg/ml in Tris-HCl is suitable. Suitable substrates include those substrates that contain a defined cross-linking chemistry (e.g., plastic substrates, such as polystyrene, styrene, or polypropylene substrates from Corning Costar Corp. (Cambridge, MA), for example) . If desired, a beaded particle, e.g., beaded agarose or beaded sepharose, can be used as the substrate.

Binding of the test compound to the polypeptides according to the invention can be detected by any of a variety of artknown methods. For example, a specific antibody can be used in an immunoassay. If desired, the antibody can be labeled (e.g., fluorescently or with a radioisotope) and detected directly (see, e.g., West and McMahon, J. Cell Biol. 74:264, 1977). Alternatively, a second antibody can be used for detection (e.g., a labeled antibody that binds the Fc portion of an anti-AN97 antibody). In an alternative detection method, the

PEC43 polypeptide is labeled, and the label is detected (e.g., by labeling a PEC43 polypeptide with a radioisotope, fluorophore, chromophore, or the like). In still another method, the PEC43 polypeptide is produced as a fusion protein with a protein that can be detected optically, e.g., green fluorescent protein (which can
5 be detected under UV light). In an alternative method, the PEC43 polypeptide can be covalently attached to or fused with an enzyme having a detectable enzymatic activity, such as horse radish peroxidase, alkaline phosphatase, α -galactosidase, or glucose oxidase. Genes encoding all of these enzymes have been cloned and are readily available for use by those of skill in the art. If desired, the fusion protein
10 can include an antigen, and such an antigen can be detected and measured with a polyclonal or monoclonal antibody using conventional methods. Suitable antigens include enzymes (e.g., horse radish peroxidase, alkaline phosphatase, and α -galactosidase) and non-enzymatic polypeptides (e.g., serum proteins, such as BSA and globulins, and milk proteins, such as caseins).

15

Epitopes, antigens and immunogens.

In another aspect, the invention provides a peptide or polypeptide comprising an epitope-bearing portion of a polypeptide of the
20 invention. The epitope of this polypeptide portion is an immunogenic or antigenic epitope of a polypeptide of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. These immunogenic epitopes are believed to be confined to a few loci on the molecule. On the other hand, a region of a protein molecule to which
25 an antibody can bind is defined as an "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes. See, for instance, Geysen, H. M. et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1984).

As to the selection of peptides or polypeptides bearing an
30 antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for instance, Sutcliffe, J. G. et al., Science 219:660-666 (1984). Peptides capable of eliciting
35 protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined

neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals. Peptides that are extremely hydrophobic and those of six or fewer residues generally are ineffective at inducing antibodies that bind to the mimicked protein; longer, soluble peptides, especially those
5 containing proline residues, usually are effective. Sutcliffe et al., supra, at 661. For instance, 18 of 20 peptides designed according to these guidelines, containing 8-39 residues covering 75% of the sequence of the influenza virus hemagglutinin HA1 polypeptide chain, induced antibodies that reacted with the HA1 protein or intact virus; and 12/12 peptides from the MuLV polymerase and 18/18 from the
10 rabies glycoprotein induced antibodies that precipitated the respective proteins.

Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. Thus, a high proportion of hybridomas obtained by fusion of spleen cells from donors immunized with an
15 antigen epitope-bearing peptide generally secrete antibody reactive with the native protein. Sutcliffe et al., supra, at 663. The antibodies raised by antigenic epitope bearing peptides or polypeptides are useful to detect the mimicked protein, and antibodies to different peptides may be used for tracking the fate of various regions of a protein precursor which undergoes posttranslation
20 processing. The peptides and anti-peptide antibodies may be used in a variety of qualitative or quantitative assays for the mimicked protein, for instance in competition assays since it has been shown that even short peptides (e.g., about 9 amino acids) can bind and displace the larger peptides in immunoprecipitation assays. See, for instance, Wilson, I.A. et al., Cell 37:767-778 at 777 (1984). The
25 anti-peptide antibodies of the invention also are useful for purification of the mimicked protein, for instance, by adsorption chromatography using methods well known in the art.

Antigenic epitope-bearing peptides and polypeptides of the invention designed according to the above guidelines preferably contain a
30 sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. However, peptides or polypeptides comprising a larger portion of an amino acid sequence of a polypeptide of the invention, containing about 30 to about 50 amino acids, or any length up to and
35 including the entire amino acid sequence of a polypeptide of the invention, also are considered epitope-bearing peptides or polypeptides of the invention and also

are useful for inducing antibodies that react with the mimicked protein. Preferably, the amino acid sequence of the epitope-bearing peptide is selected to provide substantial solubility in aqueous solvents (i.e., the sequence includes relatively hydrophilic residues and highly hydrophobic sequences are preferably avoided);
5 and sequences containing proline residues are particularly preferred.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means for making peptides or polypeptides including recombinant means using nucleic acid molecules of the invention. For instance, a short epitope-bearing amino acid sequence may be fused to a larger
10 polypeptide which acts as a carrier during recombinant production and purification, as well as during immunization to produce anti-peptide antibodies.

Epitope-bearing peptides also may be synthesized using known methods of chemical synthesis. For instance, Houghten has described a simple method for synthesis of large numbers of peptides, such as 10-20 mg of 248
15 different 13 residue peptides representing single amino acid variants of a segment of the HAI polypeptide which were prepared and characterized (by ELISA-type binding studies) in less than four weeks. Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985). This "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further described in U.S. Patent No. 4,631,211 to Houghten et
20 al. (1986). In this procedure the individual resins for the solid-phase synthesis of various peptides are contained in separate solvent-permeable packets, enabling the optimal use of the many identical repetitive steps involved in solid-phase methods.

A completely manual procedure allows 500-1000 or more
25 syntheses to be conducted simultaneously. Houghten et al., supra, at 5134.

Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art. See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F.J. et al., J. Gen. Virol. 66:2347-2354
30 (1985).

Generally, animals may be immunized with free peptide; however, anti-peptide antibody titer may be boosted by coupling of the peptide to a macromolecular carrier, such as keyhole limpet hemocyanin (KLH) or tetanus toxoid. For instance, peptides containing cysteine may be coupled to carrier using
35 a linker such as maleimidobenzoyl-N-hydroxysuccinimide ester (MBS), while other peptides may be coupled to carrier using a more general linking agent such as

glutaraldehyde.

Animals such as rabbits, rats and mice are immunized with either free or carriercoupled peptides, for instance, by intraperitoneal and/or intradermal injection of emulsions containing about 100 ug peptide or carrier
5 protein and Freund's adjuvant. Several booster injections may be needed, for instance, at intervals of about two weeks, to provide a useful titer of anti-peptide antibody which can be detected, for example, by ELISA assay using free peptide adsorbed to a solid surface. The titer of anti-peptide antibodies in serum from an immunized animal may be increased by selection of anti-peptide antibodies, for
10 instance, by adsorption to the peptide on a solid support and elution of the selected antibodies according to methods well known in the art.

Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art. For
15 instance, Geysen et al., 1984, supra, discloses a procedure for rapid concurrent synthesis on solid supports of hundreds of peptides of sufficient purity to react in an enzyme-linked immunosorbent assay. Interaction of synthesized peptides with antibodies is then easily detected without removing them from the support. In this manner a peptide bearing an immunogenic epitope of a desired protein may be
20 identified routinely by one of ordinary skill in the art. For instance, the immunologically important epitope in the coat protein of foot-and-mouth disease virus was located by Geysen et al. with a resolution of seven amino acids by synthesis of an overlapping set of all 208 possible hexapeptides covering the entire 213 amino acid sequence of the protein. Then, a complete replacement set
25 of peptides in which all 20 amino acids were substituted in turn at every position within the epitope were synthesized, and the particular amino acids conferring specificity for the reaction with antibody were determined. Thus, peptide analogs of the epitope-bearing peptides of the invention can be made routinely by this method. U.S. Patent No. 4,708,781 to Geysen (1987) further describes this
30 method of identifying a peptide bearing an immunogenic epitope of a desired protein.

Further still, U.S. Patent No. 5,194,392 to Geysen (1990) describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of
35 the epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Patent No.

4,433,092 to

Geysen (1989) describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent
5 No. 5,480,971 to Houghten, R. A. et al. (1996) on Peralkylated Oligopeptide Mixtures discloses linear C1-C7-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a peralkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs
10 of the epitope-bearing peptides of the invention also can be made routinely by these methods.

Use of PEC43 pectinases in industrial processes

15 Pectins and pectinases play an important role in the food industry. Pectinases are used in the production of fruit and vegetable juices and purees, and in the extraction of useful components, like aroma compounds, or even pharmaceuticals, from plant materials.

One of the main application of pectinases is in the production of
20 clear fruit and vegetable juice. The production of juices from fruit and vegetables typically involves, after an optional washing step, grinding, crushing or otherwise destroying the integrity of the fruit or vegetables thus obtaining a fruit or vegetable pulp. Subsequently, the pulp may be treated with enzymes to decrease soluble pectin, a process generally referred to as maceration. After maceration, the pulp is
25 ready for pressing, leading to a juice fraction and a residue fraction, the latter being referred to as the pomace. The juice obtained after pressing is usually pasteurised, optionally with recovery of the aroma, which may then be added back at the end of the process. In order to obtain a clear concentrate, the pasteurised juice is enzymatically depectinized with the aid of pectinases, optionally pre-
30 concentrated, filtrated, optionally ultra-filtrated and concentrated to obtain a clear concentrate or juice which is ready for shipping and/or blending to obtain a clear apple juice for the consumer market

PEC43 pectinases are useful in all production processes in which the integrity of the cell wall should be destroyed in order to release
35 materials from the inside of the cells, or from the cell walls.

Pectinases can also be used to destroy the connection between cells without affecting the integrity of the separate cells. When cells should remain

intact, e.g. in the preparation of potato puree PEC43 pectinases can be used to release the structural connection between the cells.

PEC43 pectinases are also able to decrease the water-binding and hence the viscosity of pectin-containing slurries, e.g. fruit or vegetable pulps, which improves the pressability/processability of the pulp.

PEC43 pectinases are able to clarify cloudy juices. Attacking the outer pectin layer of the cloud particles changes the charge of the particles, thus promoting the aggregation of the particles.

PEC43 pectinases may also be used in the retting of flax (to release the cellulose fibres from the plant tissue) and the scouring of cotton (to increase the water absorbance and dye uptake of the cotton fibres).

Plant and pectin-containing materials include plant pulp, parts of plants and plant extracts. In the context of this invention an extract from a plant material is any substance which can be derived from plant material by extraction (mechanical and/or chemical), processing or by other separation techniques. The extract may be juice, nectar, base, or concentrates made thereof. The plant material may comprise or be derived from vegetables, e.g., carrots, celery, onions, legumes or leguminous plants (soy, soybean, peas) or fruit, e.g., pome or seed fruit (apples, pears, quince etc.), grapes, tomatoes, citrus (orange, lemon, lime, mandarin), melons, prunes, cherries, black currants, redcurrants, raspberries, strawberries, cranberries, pineapple and other tropical fruits, trees and parts thereof (e.g. pollen, from pine trees).

Fruit juice that needs to be filtrated after or during processing tends to clot the filters after some time. This is due to the inability of certain enzymes to properly digest fruit polysaccharides which causes the fouling of the filters.. Use of the enzymes according to the invention prevents fouling of filters because of the unique specificity of the PEC43 pectinase.

PEC43 pectinases may conveniently be produced in microorganisms. Microbial pectinases are available from a variety of sources; Bacillus spec. are a common source of bacterial enzymes, whereas fungal enzymes are commonly produced in Aspergillus spec.

In the above processes, it is advantageous to use pectinases that are obtained by recombinant DNA techniques. Such recombinant enzymes have a number of advantages over their traditionally purified counterparts. Recombinant enzymes may be produced at a low cost price, high yield, free from contaminating agents like bacteria or viruses but also free from bacterial toxins or

contaminating other enzyme activities.

The polypeptides of the invention may be used to treat plant material including plant pulp and plant extracts. For example, they may be used to treat apple pulp and/or raw juice during the production of apple juice. They may also be used to treat liquid or solid foodstuffs or edible foodstuff ingredients. Conveniently the polypeptide of the invention is combined with suitable (solid or liquid) carriers or diluents including buffers to produce a composition or enzyme preparation. The polypeptide is typically stably formulated either in liquid or dry form. Typically, the product is made as a composition which will optionally include, for example, a stabilising buffer and/or preservative. The compositions may also include other enzymes capable of digesting plant material or pectin, for example other pectinases such as an endo-arabinanase, rhamnogalacturonases, and/or polygalacturonase. For certain applications, immobilization of the enzyme on a solid matrix or incorporation on or into solid carrier particles may be preferred. The composition may also include a variety of other plant material-degrading enzymes, for example cellulases and other pectinases.

The polypeptides and compositions of the invention may therefore be used in a method of processing plant material to degrade or modify the pectin constituents of the cell walls of the plant material

Typically, the polypeptides of the invention are used as a composition/ enzyme preparation as described above. The composition will generally be added to plant pulp obtainable by, for example mechanical processing such as crushing or milling plant material. Incubation of the composition with the plant will typically be carried out for a time of from 10 minutes to 5 hours, such as 30 minutes to 2 hours, preferably for about 1 hour. The processing temperature is preferably 10-55°C, e.g. from 15 to 25°C, optimally about 20°C and one can use 10-300g, preferably 30-70g, optimally about 50g of enzyme per ton of material to be treated. All the enzyme(s) or their compositions used may be added sequentially or at the same time to the plant pulp. Depending on the composition of the enzyme preparation the plant material may first be macerated (e.g. to a purée) or liquefied. Using the polypeptides of the invention processing parameters such as the yield of the extraction, viscosity of the extract and/or quality of the extract can be improved.

Alternatively, or in addition to the above, a polypeptide of the invention may be added to the raw juice obtained from pressing or liquefying the

plant pulp. Treatment of the raw juice may be carried out in a similar manner to the plant pulp in respect of dosage, temperature and holding time. Again, other enzymes than pectinases may be included.

5 A composition containing a polypeptide of the invention may also be used during the preparation of fruit or vegetable purees. The end product of these processes is typically heat-treated at 85°C for a time of from 1 minute to 1 hour, under conditions to partially or fully inactivate the polypeptides of the invention.

10 Due to the highly specific action on pectins the polypeptides of the invention may also be used to prepare pectins with modified characteristics, e.g. modified gelation capacities for specific applications.

The polypeptides of the invention may be added to animal feeds rich in pectin, e.g. soy-containing food, to improve the breakdown of the plant cell wall leading to improved utilisation of the plant nutrients by the animal. The polypeptides of the invention may be added to the feed or silage if pre-soaking or wet diets are preferred. Advantageously, the polypeptides of the invention may continue to degrade pectins in the feed in vivo. Fungal derived polypeptides of the invention in particular generally have lower pH optima and are capable of releasing important nutrients in such acidic environments as the stomach of an animal. The invention thus also contemplates (e.g. animal) feeds or foodstuffs comprising one or more polypeptides of the invention.

20 A PEC43 pectinase may also be advantageously used during the production of milk substitutes (or replacers) from soy bean. These milk substitutes can be consumed by both humans and animals. A typical problem during the preparation of these milk substitutes is the high viscosity of the soy bean slurry, resulting in the need for an undesirable dilution of the slurry to a concentration of dry solids of 10 to 15%. An enzyme preparation containing a polypeptide of the invention can be added to, or during the processing of, the slurry, enabling processing at a higher concentration (typically 40 to 50%) dry solids. Addition of pectinases of the invention to a soy suspension as described in WO 95/29598 results in a decrease in waterbinding and a concomitant decrease in viscosity of the slurry.

The enzyme may also be used in the preparation of savoury product(s), e.g. from soy bean.

35 The invention also relates to the use of the PEC43 pectinase according to the invention in a selected number of industrial and pharmaceutical

processes. Despite the long term experience obtained with these processes, the pectinase according to the invention features a number of significant advantages over the enzymes currently used. Depending on the specific application, these advantages can include aspects like lower production costs, higher specificity
5 towards the substrate, less antigenic, less undesirable side activities, higher yields when produced in a suitable microorganism, more suitable pH and temperature ranges, better tastes of the final product as well as food grade and kosher aspects.

An important aspect of the pectinases according to the invention
10 is that they cover a whole range of pH and temperature optima which are ideally suited for a variety of applications. For example many large scale processes benefit from relatively high processing temperatures of 50 degrees C or higher, e.g. to control the risks of microbial infections. Several pectinases according to the invention comply with this demand but at the same time they are not that heat
15 stable that they resist attempts to inactivate the enzyme by an additional heat treatment. The latter feature allows production routes that yield final products free of residual enzyme activity. Similarly many feed and food products have slightly acidic pH values so that pectinases with acidic or near neutral pH optima are preferred for their processing. A PEC43 pectinase according to the invention
20 complies with this requirement as well.

The polypeptides of the invention may also be added to animal feeds rich in pectin or xylogalacturonan, to reduce anti-nutritional effects of plant pectins and consequently improve production performances and welfare of the animals, as well reducing environmental pollution.

25 It has been shown that plant pectins increase digesta viscosity, fermentation and colonisation of bacterial proliferation in the gastro intestinal tract (Langhout, D.J. (1998). The role of the intestinal flora as affected by non-starch polysaccharides in broiler chicks. PhD thesis, Agricultural University Wageningen, The Netherlands). As a consequence, several digestive physiological parameters
30 are altered (Johnson, I.T. and Gee, J.M. (1981). Effect of gel-forming gums on the intestinal unstirred layer and sugar transport in vitro. Gut 22:398-403. Johnson, I.T. and Gee, J.M. (1986). Gastrointestinal adaption in response to soluble non-available polysaccharides in the rat. British Journal of Nutrition 55:497-505, Johnson, I.T. , Gee, J.M. and Mahoney, R.R. (1984). Effect of dietary
35 supplements of guar gum or cellulose on intestinal cell proliferation, enzyme levels and glucose transport in the rat. British Journal of Nutrition 52:477-487) resulting

in reduced dietary nutrient, mineral and energy utilization (Langhout, D.J. and Schutte, J.B. (1996). Nutritional implications of pectins in chicks in relation to esterification and origin of pectins. Poultry Science 75:1236-1242, Smits, C.M.H., Veldman, A., Verkade, H.J. and Beynen, A.C. (1998). The inhibitory effect of carboxymethylcellulose with high viscosity on lipid absorption in broiler chickens coincides with reduced bile salt concentration and raised microbial numbers in the small intestine. Poultry Science 77:1534-1539). The magnitude efficiency of dietary nutrient utilization is reduced, the effectivity in production performances (e.g. efficiency of producing milk, eggs, wool and meat) declines. Moreover, at reduced dietary nutrient utilization by the animal, excretion of and environmental pollution with e.g. nitrogenous compounds and heavy metals is increased.

Advantageous is that these fungal derived polypeptides of the invention have generally lower pH optima and their activities are therefore (more) capable of hydrolysis in such acidic environments as the stomach of an animal. With that, anti-nutritional properties of the plant cell wall fractions are already reduced early on in the digestive tract of the animal.

ABSTRACT

The invention relates to newly identified polynucleotide sequences comprising genes that encode novel pectinases isolated from *Aspergillus niger*. The invention features the full length nucleotide sequence of the novel gene, the cDNA sequence comprising the full length coding sequence of the novel pectinase as well as the amino acid sequence of the full-length functional protein and functional equivalents thereof. The invention also relates to methods of using these enzymes in industrial processes and methods of diagnosing fungal infections. Also included in the invention are cells transformed with a polynucleotide according to the invention and cells wherein an pectinase according to the invention is genetically modified to enhance or reduce its activity and/or level of expression.

CLAIMS

- 1) An isolated polynucleotide hybridisable to a polynucleotide according to SEQ ID NO:1 or SEQ ID NO: 2
- 2) An isolated polynucleotide according to claim 1 hybridisable under high stringency conditions to a polynucleotide according to SEQ ID NO:1 or SEQ ID NO: 2
- 3) An isolated polynucleotide according to claims 1 or 2 obtainable from a filamentous fungus.
- 4) An isolated polynucleotide according to claim 3 obtainable from *A. niger*.
- 5) An isolated polynucleotide encoding a polypeptide comprising an amino acid sequence according to SEQ ID NO: 3 or functional equivalents thereof.
- 6) An isolated polynucleotide encoding at least one functional domain of a polypeptide according to SEQ ID NO: 3 or functional equivalents thereof.
- 7) An isolated polynucleotide comprising a nucleotide sequence according to SEQ ID NO: 1 or SEQ ID NO: 2 or functional equivalents thereof
- 8) An isolated polynucleotide according to SEQ ID NO: 1 or SEQ ID NO: 2.
- 9) A vector comprising a polynucleotide sequence according to claims 1 to 8.
- 10) A vector according to claim 9 wherein said polynucleotide sequence according to claims 1 to 8 is operatively linked with regulatory sequences suitable for expression of said polynucleotide sequence in a suitable host cell.
- 11) A vector according to claim 10 wherein said suitable host cell is a filamentous fungus
- 12) A method for manufacturing a polynucleotide according to claims 1 – 8 or a vector according to claims 9 to 11 comprising the steps of culturing a host cell transformed with said polynucleotide or said vector and isolating said polynucleotide or said vector from said host cell.
- 13) An isolated polypeptide according to SEQ ID NO: 3 or functional equivalents thereof.
- 14) An isolated polypeptide according to claim 13 obtainable from *Aspergillus niger*
- 15) An isolated polypeptide obtainable by expressing a polynucleotide according to claims 1 to 8 or a vector according to claims 9 to 11 in an appropriate host cell, e.g. *Aspergillus niger*.
- 16) Recombinant pectinase comprising a functional domain of a PEC43 polypeptide.

- 17) A method for manufacturing a polypeptide according to claims 13 to 16 comprising the steps of transforming a suitable host cell with an isolated polynucleotide according to claims 1 to 8 or a vector according to claims 9 to 11, culturing said cell under conditions allowing expression of said polynucleotide and optionally purifying the encoded polypeptide from said cell or culture medium.
- 18) A recombinant host cell comprising a polynucleotide according to claims 1 to 8 or a vector according to claims 9 to 11.
- 19) A recombinant host cell expressing a polypeptide according to claims 13 to 16.
- 20) Purified antibodies reactive with a polypeptide according to claims 13 to 16.
- 21) Fusion protein comprising a polypeptide sequence according to claims 13 to 16.

SEQUENCE LISTING

<110> DSM NV.

<120> Novel pectinase PEC43 and uses thereof

<130> 20802EP/P0

<160> 3

<170> XML to WIPO ST.25 Converter - <http://www.biomax.de/>

<210> 1

<211> 3446

<212> DNA

<213> *Aspergillus niger*

<400> 1

gagtgcagaca	tgtgcagcga	taaataaaca	tgctgtacac	tagcggctag	tataaacatc	60
tccccttcga	ttttatacgg	aacgtgagtt	attgtacaat	ggaacaatag	actgtttgta	120
taatgccaat	ctactctttt	ctctacttct	cagcatcgcc	ttttggagcg	ggccattgca	180
aagacaaaaa	gtgggatagt	tagtcatatc	tataggcctc	ataccagaac	tattataact	240
cgtctagcca	ccgtgccggg	ttgaatttgc	cacggttaag	atgcagagcg	tcgctaaaat	300
tgacttcggt	tgagtctcag	cagacaaaacg	tgttccccgg	gatggcctca	tcacttcccg	360
ggatagggat	cggtttgcaa	gagaggggtca	actaggctat	agagaaaaga	tagaagttat	420
agttacagtt	aggctcaata	cctcatgtcc	atthttggtga	gtgcttaggg	tgttacactg	480
aacttcatta	attcggtcgc	aacccatgac	ctcgtacgcc	taaacactgt	gccaggaacc	540
ggccgcagac	aaccccgcaa	ccccaacccc	tgacgacctc	cacttgaaca	acaaagttct	600
tctttatata	accaagaaca	cgttggtcgg	aattgacatc	atgattaggg	aggtggtacc	660
agatthtatcg	catctgtccg	acatccgcac	cgcggaccat	caacatgcat	gatgtcatca	720
cgggaagaccg	ggggacaccg	atgcgacgct	gctaacccta	ccaacaaaag	acttcaatac	780
acctacattg	ttgtcctttt	ttcgctatth	agaaagcctc	attgagcatg	ctgacgagga	840
agctagatct	gccacgaaat	gatagcggtc	gcaaaatggc	tatacgttgc	tgtgacggcg	900
tattttocaca	agttgaggcc	ggttagccat	gattatcctg	ctctggcagc	aagggataat	960
ggtacaaaacg	gcagcgtggc	gtcgccaata	acattgactg	ttgaagagtc	gggcgggaat	1020
cagtctagtc	ctctgcttta	tgggggttatg	tttgaggtaa	gtgctgcgag	gcagcatggg	1080
attgtcctag	taacattctc	ttaggaaatg	gaccactcag	gtactatcta	gttcagtagc	1140
tctgttgaca	gtatthttgct	gagtatatth	aggtgatgga	ggcattcacg	ggcaattgct	1200
ccagaataat	ggcttccagg	gggacgatcc	cggcttgact	gcatacaaag	cggtcggggc	1260
cgtcgacctc	atgcaggact	tgataaaacc	ggtcagcggg	gcaataacgt	cttctttgca	1320
ggtatctgta	gattthtgaag	caacgggggt	tgtaggatth	gccaatacac	ggtacagtgg	1380
gatacctgtg	atgaatgcaa	cgtattcgtg	ccagttctgg	atgatgggtg	aatactcagg	1440
aacgattatg	ctacagcttg	ctggatcaac	gaatgatact	atthtatgctt	cgcataatat	1500
taccgtgaag	agctcatggg	gaaagttcac	ccattatcaa	acttctthta	attcgagcgc	1560
agctcctgat	ggtgacaatg	agtggcgctt	gctgttcaat	ggatcaaaga	tggctggagg	1620
gatgctgaac	ttcgggtctgg	tgacgtgtth	tccgcctacg	tacaagtcga	ggtacgttac	1680
attaagatta	ccaatgtcat	tctaatcggg	aacagggtcaa	atggactacg	gaatgatgtt	1740
gcgacattth	tggaaaagac	agctccctct	tttctccgat	tcccgggtgg	caataatctg	1800
taggtaatca	ccctgaaaat	gaagatcctc	taacaagcgg	atagagaagg	actacagatc	1860
gacagccgth	ggcagtgga	cctaactgth	gggcccgcgc	ttgatcgtcc	cgggagacag	1920
ggtagggcac	caccaccgag	accatgcatg	acaaagacta	aatgatgcat	aggcgactgg	1980
ttctacccaa	acaccgacgc	gctcggtaag	tctgtgcat	ccttctgccc	cacgagcatg	2040
tggtaacaca	cacagatttg	gacgaatatc	tctggtggtg	cgaagacatg	aacatggagc	2100
ccgtcctcgc	cgtctgggac	ggcaaatcat	acggcggcat	tctatccggc	gacgatctcc	2160
agccatatat	agacgatatc	atgaacgagc	ttgaggtata	cctccatgga	ttactcttat	2220
catgccgagc	tgaccgtaaa	gtacctcctc	gggcccgtca	actccacctc	cgggagtagt	2280

cgagcgcaaa	acggggcgag	caaaccatgg	tccataaact	acatcgagat	cgggaaacgaa	2340
gatgacttta	ccggcggatg	cgacacatac	ccagaccgct	tctatcagat	ttacaacgcc	2400
atcagcaaca	gctacccgaa	tatcactctc	atcgcatcga	atatagacta	cctgtgtctc	2460
ccgattgaac	ccccaccggg	cctgatatac	gactaccatt	actatcgcaa	accggatgac	2520
ctcgtcgcca	tgttcgacta	ctgggataac	cagcctcgca	cgcagcccat	catggtaggg	2580
gaatatggat	gtcgggatac	cagcgaggca	gacgggatct	tctggtcctc	tatgcaatgt	2640
agctgcagtg	aggcagcgca	tatgattggc	ctggagagga	actcggatgt	agtgaaaatg	2700
gctgcgtatg	cgccactatt	gcagcatttc	gggtatacgc	agtggtcggt	gagtttggcc	2760
tccttaccctc	cattcttgta	tgtccattat	gttagagcta	acaagaacaa	agccaacgct	2820
attcggcttc	gactctagtc	ctgactcgct	gactccctcg	acgtcatact	atgttcaacg	2880
gatgttttct	acaaatcgcg	gggatactat	cctccctgtg	aatacaactg	ccacatttgg	2940
accgctgtac	tgggttgcat	ccaggacgaa	tagcacgtat	tttgtgaagt	tggccaacta	3000
cggcgcccg	aaccagactg	tacgtgtaaa	ggtgccccaa	accaagaccg	ggcatgtaga	3060
aatgctgtat	gggccgcaaa	atgcgaccaa	tttgggtgcat	aataattactg	ttcagccgac	3120
ggtgaagaat	gtgactagct	ctagggggcat	ttattccctg	gatatgccgc	cgtgggggggt	3180
tgctgttttg	tccgtgtggt	gaggggtggt	gctgctgtat	aacaggagtt	ttattgtgga	3240
tcaagtatct	acgagtcatg	ttgtttacca	tatagttcaa	tacaaaaagg	ggaaatggaa	3300
tttaaagcga	aggaccgatg	tcgattaaga	attgaataat	cgtcgtttta	atgtatatac	3360
tacctacaac	tcgctttctc	atcgccctcc	ttcgtctgcc	gcctgctgaa	taatatacag	3420
ctaaaagtca	ccagtgactt	atgggc				3446

<210> 2

<211> 1941

<212> DNA

<213> *Aspergillus niger*

<220>

<221> CDS

<222> (1)..(1941)

<400> 2

atg	ata	gcg	gtc	gca	aaa	tgg	cta	tac	gtt	gct	gtg	acg	gcg	tat	ttc	48
Met	Ile	Ala	Val	Ala	Lys	Trp	Leu	Tyr	Val	Ala	Val	Thr	Ala	Tyr	Phe	
1			5						10					15		
cac	aag	ttg	agg	ccg	gtt	agc	cat	gat	tat	cct	gct	ctg	gca	gca	agg	96
His	Lys	Leu	Arg	Pro	Val	Ser	His	Asp	Tyr	Pro	Ala	Leu	Ala	Ala	Arg	
			20					25					30			
gat	aat	ggt	aca	aac	ggc	agc	gtg	gcg	tcg	cca	ata	aca	ttg	act	gtt	144
Asp	Asn	Gly	Thr	Asn	Gly	Ser	Val	Ala	Ser	Pro	Ile	Thr	Leu	Thr	Val	
		35					40					45				
gaa	gag	tcg	ggc	ggg	aat	cag	tct	agt	cct	ctg	ctt	tat	ggg	gtt	atg	192
Glu	Glu	Ser	Gly	Gly	Asn	Gln	Ser	Ser	Pro	Leu	Leu	Tyr	Gly	Val	Met	
		50				55					60					
ttt	gag	gta	ata	tat	tca	ggt	gat	gga	ggc	att	cac	ggg	caa	ttg	ctc	240
Phe	Glu	Val	Ile	Tyr	Ser	Gly	Asp	Gly	Gly	Ile	His	Gly	Gln	Leu	Leu	
65					70				75					80		
cag	aat	aat	ggc	ttc	cag	ggg	gac	gat	ccc	ggc	ttg	act	gca	tac	aaa	288
Gln	Asn	Asn	Gly	Phe	Gln	Gly	Asp	Asp	Pro	Gly	Leu	Thr	Ala	Tyr	Lys	
			85					90						95		
gcg	gtc	ggg	ccc	gtc	gac	ctc	atg	cag	gac	ttg	ata	aac	ccg	gtc	agc	336
Ala	Val	Gly	Pro	Val	Asp	Leu	Met	Gln	Asp	Leu	Ile	Asn	Pro	Val	Ser	
			100					105						110		
ggt	gca	ata	acg	tct	tct	ttg	cag	gta	tct	gta	gat	ttt	gaa	gca	acg	384
Gly	Ala	Ile	Thr	Ser	Ser	Leu	Gln	Val	Ser	Val	Asp	Phe	Glu	Ala	Thr	
		115					120					125				
ggg	ttt	gta	gga	ttt	gcc	aat	aca	ggg	tac	agt	ggg	ata	cct	gtg	atg	432

Gly	Phe	Val	Gly	Phe	Ala	Asn	Thr	Gly	Tyr	Ser	Gly	Ile	Pro	Val	Met	
130						135					140					
aat	gca	acg	tat	tcg	tgc	cag	ttc	tgg	atg	atg	ggt	gaa	tac	tca	gga	480
Asn	Ala	Thr	Tyr	Ser	Cys	Gln	Phe	Trp	Met	Met	Gly	Glu	Tyr	Ser	Gly	
145					150					155					160	
acg	att	atg	cta	cag	ctt	gct	gga	tca	acg	aat	gat	act	att	tat	gct	528
Thr	Ile	Met	Leu	Gln	Leu	Ala	Gly	Ser	Thr	Asn	Asp	Thr	Ile	Tyr	Ala	
				165					170						175	
tcg	cat	aat	att	acc	gtg	aag	agc	tca	tgg	gga	aag	ttc	acc	cat	tat	576
Ser	His	Asn	Ile	Thr	Val	Lys	Ser	Ser	Trp	Gly	Lys	Phe	Thr	His	Tyr	
		180						185						190		
caa	act	tct	ttt	aat	tcg	agc	gca	gct	cct	gat	ggt	gac	aat	gag	tgg	624
Gln	Thr	Ser	Phe	Asn	Ser	Ser	Ala	Ala	Pro	Asp	Gly	Asp	Asn	Glu	Trp	
	195						200					205				
cgc	ctg	ctg	ttc	aat	gga	tca	aag	atg	gct	gga	ggg	atg	ctg	aac	ttc	672
Arg	Leu	Leu	Phe	Asn	Gly	Ser	Lys	Met	Ala	Gly	Gly	Met	Leu	Asn	Phe	
	210					215					220					
ggt	ctg	gtg	cag	ctg	ttt	ccg	cct	acg	tac	aag	tcg	agg	tca	aat	gga	720
Gly	Leu	Val	Gln	Leu	Phe	Pro	Pro	Thr	Tyr	Lys	Ser	Arg	Ser	Asn	Gly	
225					230					235					240	
cta	cgg	aat	gat	gtt	gcg	aca	ttt	ctg	gaa	aag	aca	gct	ccc	tct	ttt	768
Leu	Arg	Asn	Asp	Val	Ala	Thr	Phe	Leu	Glu	Lys	Thr	Ala	Pro	Ser	Phe	
				245					250						255	
ctc	cga	ttc	ccg	ggt	ggc	aat	aat	cta	gaa	gga	cta	cag	atc	gac	agc	816
Leu	Arg	Phe	Pro	Gly	Gly	Asn	Asn	Leu	Glu	Gly	Leu	Gln	Ile	Asp	Ser	
			260					265					270			
cgt	tgg	cag	tgg	aac	cta	act	gtt	ggg	ccc	gtc	gtt	gat	cgt	ccc	ggg	864
Arg	Trp	Gln	Trp	Asn	Leu	Thr	Val	Gly	Pro	Val	Val	Asp	Arg	Pro	Gly	
	275					280						285				
aga	cag	ggc	gac	tgg	ttc	tac	cca	aac	acc	gac	gcg	ctc	gga	ttg	gac	912
Arg	Gln	Gly	Asp	Trp	Phe	Tyr	Pro	Asn	Thr	Asp	Ala	Leu	Gly	Leu	Asp	
	290					295				300						
gaa	tat	ctc	tgg	tgg	tgc	gaa	gac	atg	aac	atg	gag	ccc	gtc	ctc	gcc	960
Glu	Tyr	Leu	Trp	Trp	Cys	Glu	Asp	Met	Asn	Met	Glu	Pro	Val	Leu	Ala	
305					310				315						320	
gtc	tgg	gac	ggc	aaa	tca	tac	ggc	ggc	att	cta	tcc	ggc	gac	gat	ctc	1008
Val	Trp	Asp	Gly	Lys	Ser	Tyr	Gly	Gly	Ile	Leu	Ser	Gly	Asp	Asp	Leu	
				325				330							335	
cag	cca	tat	ata	gac	gat	atc	atg	aac	gag	ctt	gag	tac	ctc	ctc	ggg	1056
Gln	Pro	Tyr	Ile	Asp	Asp	Ile	Met	Asn	Glu	Leu	Glu	Tyr	Leu	Leu	Gly	
			340					345					350			
ccg	gtc	aac	tcc	acc	tac	ggg	agt	atg	cga	gcg	caa	aac	ggg	cgc	agc	1104
Pro	Val	Asn	Ser	Thr	Tyr	Gly	Ser	Met	Arg	Ala	Gln	Asn	Gly	Arg	Ser	
		355					360					365				
aaa	cca	tgg	tcc	ata	aac	tac	atc	gag	atc	ggg	aac	gaa	gat	gac	ttt	1152
Lys	Pro	Trp	Ser	Ile	Asn	Tyr	Ile	Glu	Ile	Gly	Asn	Glu	Asp	Asp	Phe	
	370				375					380						
acc	ggc	gga	tgc	gac	aca	tac	cca	gac	cgc	ttc	tat	cag	att	tac	aac	1200
Thr	Gly	Gly	Cys	Asp	Thr	Tyr	Pro	Asp	Arg	Phe	Tyr	Gln	Ile	Tyr	Asn	
385					390				395						400	
gcc	atc	agc	aac	agc	tac	ccg	aat	atc	act	ctc	atc	gca	tcg	aat	ata	1248
Ala	Ile	Ser	Asn	Ser	Tyr	Pro	Asn	Ile	Thr	Leu	Ile	Ala	Ser	Asn	Ile	
				405				410							415	
gac	tac	ctg	tgt	ctc	ccg	att	gaa	ccc	cca	ccc	ggc	ctg	ata	tac	gac	1296
Asp	Tyr	Leu	Cys	Leu	Pro	Ile	Glu	Pro	Pro	Pro	Gly	Leu	Ile	Tyr	Asp	
			420				425						430			
tac	cat	tac	tat	cgc	aaa	ccg	gat	gac	ctc	gtc	gcc	atg	ttc	gac	tac	1344

Tyr	His	Tyr	Tyr	Arg	Lys	Pro	Asp	Asp	Leu	Val	Ala	Met	Phe	Asp	Tyr	
		435					440					445				
tgg	gat	aac	cag	cct	cgc	acg	cag	ccc	atc	atg	gta	ggg	gaa	tat	gga	1392
Trp	Asp	Asn	Gln	Pro	Arg	Thr	Gln	Pro	Ile	Met	Val	Gly	Glu	Tyr	Gly	
		450					455				460					
tgt	cgg	gat	acc	agc	gag	gca	gac	ggg	atc	ttc	tgg	tcc	tct	atg	caa	1440
Cys	Arg	Asp	Thr	Ser	Glu	Ala	Asp	Gly	Ile	Phe	Trp	Ser	Ser	Met	Gln	
465					470					475					480	
tgt	agc	tgc	agt	gag	gca	gcg	cat	atg	att	ggc	ctg	gag	agg	aac	tcg	1488
Cys	Ser	Cys	Ser	Glu	Ala	Ala	His	Met	Ile	Gly	Leu	Glu	Arg	Asn	Ser	
				485					490					495		
gat	gta	gtg	aaa	atg	gct	gcg	tat	gcg	cca	cta	ttg	cag	cat	ttc	ggg	1536
Asp	Val	Val	Lys	Met	Ala	Ala	Tyr	Ala	Pro	Leu	Leu	Gln	His	Phe	Gly	
			500					505						510		
tat	acg	cag	tgg	tcg	cca	acg	cta	ttc	ggc	ttc	gac	tct	agt	cct	gac	1584
Tyr	Thr	Gln	Trp	Ser	Pro	Thr	Leu	Phe	Gly	Phe	Asp	Ser	Ser	Pro	Asp	
		515					520					525				
tcg	ctg	act	ccc	tcg	acg	tca	tac	tat	gtt	caa	cgg	atg	ttt	tct	aca	1632
Ser	Leu	Thr	Pro	Ser	Thr	Ser	Tyr	Tyr	Val	Gln	Arg	Met	Phe	Ser	Thr	
		530					535					540				
aat	cgc	ggg	gat	act	atc	ctc	cct	gtg	aat	aca	act	gcc	aca	ttt	gga	1680
Asn	Arg	Gly	Asp	Thr	Ile	Leu	Pro	Val	Asn	Thr	Thr	Ala	Thr	Phe	Gly	
545					550					555					560	
ccg	ctg	tac	tgg	gtt	gca	tcc	agg	acg	aat	agc	acg	tat	ttt	gtg	aag	1728
Pro	Leu	Tyr	Trp	Val	Ala	Ser	Arg	Thr	Asn	Ser	Thr	Tyr	Phe	Val	Lys	
				565					570					575		
ttg	gcc	aac	tac	ggc	gcc	cag	aac	cag	act	gta	cgt	gta	aag	gtg	ccc	1776
Leu	Ala	Asn	Tyr	Gly	Ala	Gln	Asn	Gln	Thr	Val	Arg	Val	Lys	Val	Pro	
		580						585					590			
caa	acc	aag	acc	ggg	cat	gta	gaa	atg	ctg	tat	ggg	ccg	caa	aat	gcg	1824
Gln	Thr	Lys	Thr	Gly	His	Val	Glu	Met	Leu	Tyr	Gly	Pro	Gln	Asn	Ala	
		595					600					605				
acc	aat	ttg	gtg	cat	aat	att	act	gtt	cag	ccg	acg	gtg	aag	aat	gtg	1872
Thr	Asn	Leu	Val	His	Asn	Ile	Thr	Val	Gln	Pro	Thr	Val	Lys	Asn	Val	
		610				615						620				
act	agc	tct	agg	ggc	att	tat	tcc	ctg	gat	atg	ccg	ccg	tgg	ggg	gtt	1920
Thr	Ser	Ser	Arg	Gly	Ile	Tyr	Ser	Leu	Asp	Met	Pro	Pro	Trp	Gly	Val	
625					630					635					640	
gct	gtt	ttg	tcg	gtg	tgg	tga										1941
Ala	Val	Leu	Ser	Val	Trp											
				645												

<210> 3

<211> 646

<212> PRT

<213> *Aspergillus niger*

<400> 3

Met	Ile	Ala	Val	Ala	Lys	Trp	Leu	Tyr	Val	Ala	Val	Thr	Ala	Tyr	Phe	
1				5					10					15		
His	Lys	Leu	Arg	Pro	Val	Ser	His	Asp	Tyr	Pro	Ala	Leu	Ala	Ala	Arg	
			20					25					30			
Asp	Asn	Gly	Thr	Asn	Gly	Ser	Val	Ala	Ser	Pro	Ile	Thr	Leu	Thr	Val	
		35					40					45				
Glu	Glu	Ser	Gly	Gly	Asn	Gln	Ser	Ser	Pro	Leu	Leu	Tyr	Gly	Val	Met	
	50					55					60					

Phe	Glu	Val	Ile	Tyr	Ser	Gly	Asp	Gly	Gly	Ile	His	Gly	Gln	Leu	Leu
65					70					75					80
Gln	Asn	Asn	Gly	Phe	Gln	Gly	Asp	Asp	Pro	Gly	Leu	Thr	Ala	Tyr	Lys
				85					90					95	
Ala	Val	Gly	Pro	Val	Asp	Leu	Met	Gln	Asp	Leu	Ile	Asn	Pro	Val	Ser
			100					105					110		
Gly	Ala	Ile	Thr	Ser	Ser	Leu	Gln	Val	Ser	Val	Asp	Phe	Glu	Ala	Thr
		115					120					125			
Gly	Phe	Val	Gly	Phe	Ala	Asn	Thr	Gly	Tyr	Ser	Gly	Ile	Pro	Val	Met
	130					135					140				
Asn	Ala	Thr	Tyr	Ser	Cys	Gln	Phe	Trp	Met	Met	Gly	Glu	Tyr	Ser	Gly
145					150					155					160
Thr	Ile	Met	Leu	Gln	Leu	Ala	Gly	Ser	Thr	Asn	Asp	Thr	Ile	Tyr	Ala
				165					170					175	
Ser	His	Asn	Ile	Thr	Val	Lys	Ser	Ser	Trp	Gly	Lys	Phe	Thr	His	Tyr
			180					185					190		
Gln	Thr	Ser	Phe	Asn	Ser	Ser	Ala	Ala	Pro	Asp	Gly	Asp	Asn	Glu	Trp
		195					200					205			
Arg	Leu	Leu	Phe	Asn	Gly	Ser	Lys	Met	Ala	Gly	Gly	Met	Leu	Asn	Phe
	210					215						220			
Gly	Leu	Val	Gln	Leu	Phe	Pro	Pro	Thr	Tyr	Lys	Ser	Arg	Ser	Asn	Gly
225					230					235					240
Leu	Arg	Asn	Asp	Val	Ala	Thr	Phe	Leu	Glu	Lys	Thr	Ala	Pro	Ser	Phe
				245					250					255	
Leu	Arg	Phe	Pro	Gly	Gly	Asn	Asn	Leu	Glu	Gly	Leu	Gln	Ile	Asp	Ser
			260					265					270		
Arg	Trp	Gln	Trp	Asn	Leu	Thr	Val	Gly	Pro	Val	Val	Asp	Arg	Pro	Gly
		275					280					285			
Arg	Gln	Gly	Asp	Trp	Phe	Tyr	Pro	Asn	Thr	Asp	Ala	Leu	Gly	Leu	Asp
	290					295					300				
Glu	Tyr	Leu	Trp	Trp	Cys	Glu	Asp	Met	Asn	Met	Glu	Pro	Val	Leu	Ala
305					310					315					320
Val	Trp	Asp	Gly	Lys	Ser	Tyr	Gly	Gly	Ile	Leu	Ser	Gly	Asp	Asp	Leu
				325					330					335	
Gln	Pro	Tyr	Ile	Asp	Asp	Ile	Met	Asn	Glu	Leu	Glu	Tyr	Leu	Leu	Gly
			340					345					350		
Pro	Val	Asn	Ser	Thr	Tyr	Gly	Ser	Met	Arg	Ala	Gln	Asn	Gly	Arg	Ser
		355					360					365			
Lys	Pro	Trp	Ser	Ile	Asn	Tyr	Ile	Glu	Ile	Gly	Asn	Glu	Asp	Asp	Phe
	370					375						380			
Thr	Gly	Gly	Cys	Asp	Thr	Tyr	Pro	Asp	Arg	Phe	Tyr	Gln	Ile	Tyr	Asn
385					390					395					400
Ala	Ile	Ser	Asn	Ser	Tyr	Pro	Asn	Ile	Thr	Leu	Ile	Ala	Ser	Asn	Ile
				405					410					415	
Asp	Tyr	Leu	Cys	Leu	Pro	Ile	Glu	Pro	Pro	Pro	Gly	Leu	Ile	Tyr	Asp
			420					425					430		
Tyr	His	Tyr	Tyr	Arg	Lys	Pro	Asp	Asp	Leu	Val	Ala	Met	Phe	Asp	Tyr
		435					440					445			
Trp	Asp	Asn	Gln	Pro	Arg	Thr	Gln	Pro	Ile	Met	Val	Gly	Glu	Tyr	Gly
	450					455					460				
Cys	Arg	Asp	Thr	Ser	Glu	Ala	Asp	Gly	Ile	Phe	Trp	Ser	Ser	Met	Gln
465					470					475					480
Cys	Ser	Cys	Ser	Glu	Ala	Ala	His	Met	Ile	Gly	Leu	Glu	Arg	Asn	Ser
				485					490					495	
Asp	Val	Val	Lys	Met	Ala	Ala	Tyr	Ala	Pro	Leu	Leu	Gln	His	Phe	Gly
			500					505					510		
Tyr	Thr	Gln	Trp	Ser	Pro	Thr	Leu	Phe	Gly	Phe	Asp	Ser	Ser	Pro	Asp

[illegible]

PCT/EP2003/005726

